approximately three other loan records was also included in the sample. It is assumed that these houses are not energy efficient and are considered part of the "control" group. Furthermore, the sample is restricted to single-family owner-occupied houses whose loans originated from January 2002 and loans that were used for purchase only.

All the loan level variables including payment stream were provided by CoreLogic. Prepayment is defined as loans being paid off prematurely. Consistent with prior work, 90 days delinquency is used to capture mortgage default. Similarly, the key risk determinants at origination are included in the model: borrower credit score (FICO), loan-to-value ratio, loan type (conventional/government and non-profit organizations backed), local unemployment rate, neighborhood income, house value relative to the area median value, debt-to-income ratio, size of the house, and age of the house (Table 1).

[Insert Table 1]

We constructed the neighborhood level variables from multiple sources. CoreLogic's MarketTrends database was used to include variables such as the number of foreclosures and the number of properties own by lenders (REOs) in the area. Unemployment rate, median housing value, and household income were retrieved from the 2006-2010 Census American Community Survey at the census-tract level. Geographic weighting was used to aggregate the data to the zip-code level.

In addition to the Energy Star and HERS rating, we included a number of other energy-use-related variables in the analysis. These include: number of cooling degree days, number of heating degree days, electricity prices, and area of the house. Weather data, such as average annual (over the last decade) cooling degree days and heating degree days, are obtained from the National Climatic Data Center. Each weather station is assigned to a block group and then data were aggregated to zip-code level through geographic weighting. Electricity prices are used as a proxy